

Machine-Learning-Method-Based Inversion of Shallow Bathymetric Maps Using ICESat-2 ATL03 Data

Tao Xie , Ruiyao Kong, Abdul Nurunnabi , Shuying Bai, and Xuehong Zhang

Abstract—The application of empirical methods for satellite-derived bathymetry is limited by the lack of in situ bathymetric data in remote, inaccessible areas. This challenge has been addressed with the launch of Ice, Cloud, and land Elevation Satellite-2 (ICESat-2). This study provides an accurate bathymetric photon extraction process for ICESat-2 ATL03 data, and the R^2 value of the bathymetric photons obtained using this process and airborne bathymetric LIDAR data is up to 99%. Next, based on two types of remote sensing data, ICESat-2 and Sentinel-2, machine learning models, including linear regression (LR), light gradient boosting machine (LightGBM), and categorical boosting (CatBoost), were trained to obtain bathymetric maps. The experimental results show that the mean root mean square error (RMSE), mean absolute error (MAE), and mean relative error (MRE) values of the LR models are less than 3.02 m, 2.38 m, and 86.03%, respectively. The mean RMSE, MAE, and MRE values of the LightGBM and CatBoost models are less than 0.91 m, 0.66 m, and 23.17%, respectively. It is concluded that the proposed denoising process for ICESat-2 ATL03 data is effective, and the results of the bathymetric maps obtained using these data are satisfactory. Thus, the proposed approach is effective, and this strategy can be used to replace conventional bathymetric inversion methods to obtain high-accuracy bathymetric maps.

Index Terms—Bathymetry, denoising, ice, Cloud, and land elevation satellite-2 (ICESat-2), machine learning (ML), sentinel-2.

Manuscript received 7 November 2022; revised 6 January 2023, 23 February 2023, and 15 March 2023; accepted 17 March 2023. Date of publication 23 March 2023; date of current version 17 April 2023. This work was supported in part by the National Key R&D Program of China under Grants 2022YFC3004200 and 2022YFC3104900, in part by the National Natural Science Foundation of China under Grant 42176180, and in part by the Natural Science Foundation of Jiangsu Province under Grants YJGL-YF-2020-16 and JSZRHYKJ202114. The work of Abdul Nurunnabi was supported as part of Project 2019-05-030-24, SOLSTICE - Programme Fonds Européen de Développement Régional (FEDER)/Ministère de l'Économie of the G. D. of Luxembourg. (*Corresponding author: Tao Xie.*)

Tao Xie is with the School of Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China, also with the Laboratory for Regional Oceanography and Numerical Modeling, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266000, China, also with the Technology Innovation Center for Integration Applications in Remote Sensing and Navigation, Ministry of Natural Resources, Nanjing, Jiangsu 210044, China, and also with the Jiangsu Province Engineering Research Center of Collaborative Navigation/Positioning and Smart Application, Nanjing, Jiangsu 210044, China (e-mail: xietao@nuist.edu.cn).

Ruiyao Kong, Shuying Bai, and Xuehong Zhang are with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: 20201248050@nuist.edu.cn; baishu-ying@163.com; zxhbnu@126.com).

Abdul Nurunnabi is with the Department of Geodesy and Geospatial Engineering, Institute of Civil and Environmental Engineering, University of Luxembourg, 4365 Esch-sur-Alzette, Luxembourg (e-mail: abdul.nurunnabi@uni.lu).

Digital Object Identifier 10.1109/JSTARS.2023.3260831

I. INTRODUCTION

IT IS fundamental to obtain accurate shallow water depth information for marine ecological environment research and marine safety navigation. In traditional bathymetry, bathymetry rods and echo sounders are utilized; although these tools can be used to obtain measurements with high accuracy, there are several factors and environmental constraints on measurements, and it is arduous to work in remote and unreachable areas [1]. Satellite-derived bathymetry (SDB), which has gained significant attention from researchers as a result of the advancement of remote sensing technologies, has been used to obtain bathymetric maps [1], [2], [3], [4], [5].

The current methods to derive shallow water bathymetry can be categorized mainly into theoretical [3], [4], [6], [7], semiempirical [2], [8], [9], [10], [11], and empirical [12], [13], [14] methods. The radiative transfer process serves as the foundation for the theoretical methods used to create the water depth inversion equation, which makes it challenging to obtain and calculate many of the optical parameters of water. Empirical methods are based on the relationship between the reflectance of remote sensing imagery and in situ bathymetric data. Machine learning (ML), such as random forest, is also a special form of empirical method that has high accuracy in a specific area [12], [13], [14], [15], [16], [17], [18]. Using a support vector machine, bathymetric inversion of Sint Maarten Island and Ameland Inlet was performed; the performance for shallow water depths was improved, and accurate bathymetric maps were effectively obtained [19]. However, the lack of in situ bathymetric data for constructing a method in some study areas has become an important issue limiting its application. Nonetheless, Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2) was launched successfully in 2018, offering a fresh approach to this issue [20], [21], [22], [23].

ICESat-2 carries an instrument called ATLAS (advanced topographic laser altimeter system), a highly sensitive photon-counting LiDAR system. ICESat-2 has three pairs of beams [24]. Each beam pair contains two beams, i.e., a strong beam and a weak beam, and the strong beam has four times the energy of the weak beam. Because ICESat-2's detectors are so sensitive, and for other reasons, the raw photons contain a large number of noise photons, especially during daylight hours. Consequently, there are numerous ways to extract signal photons from ATL03 data.

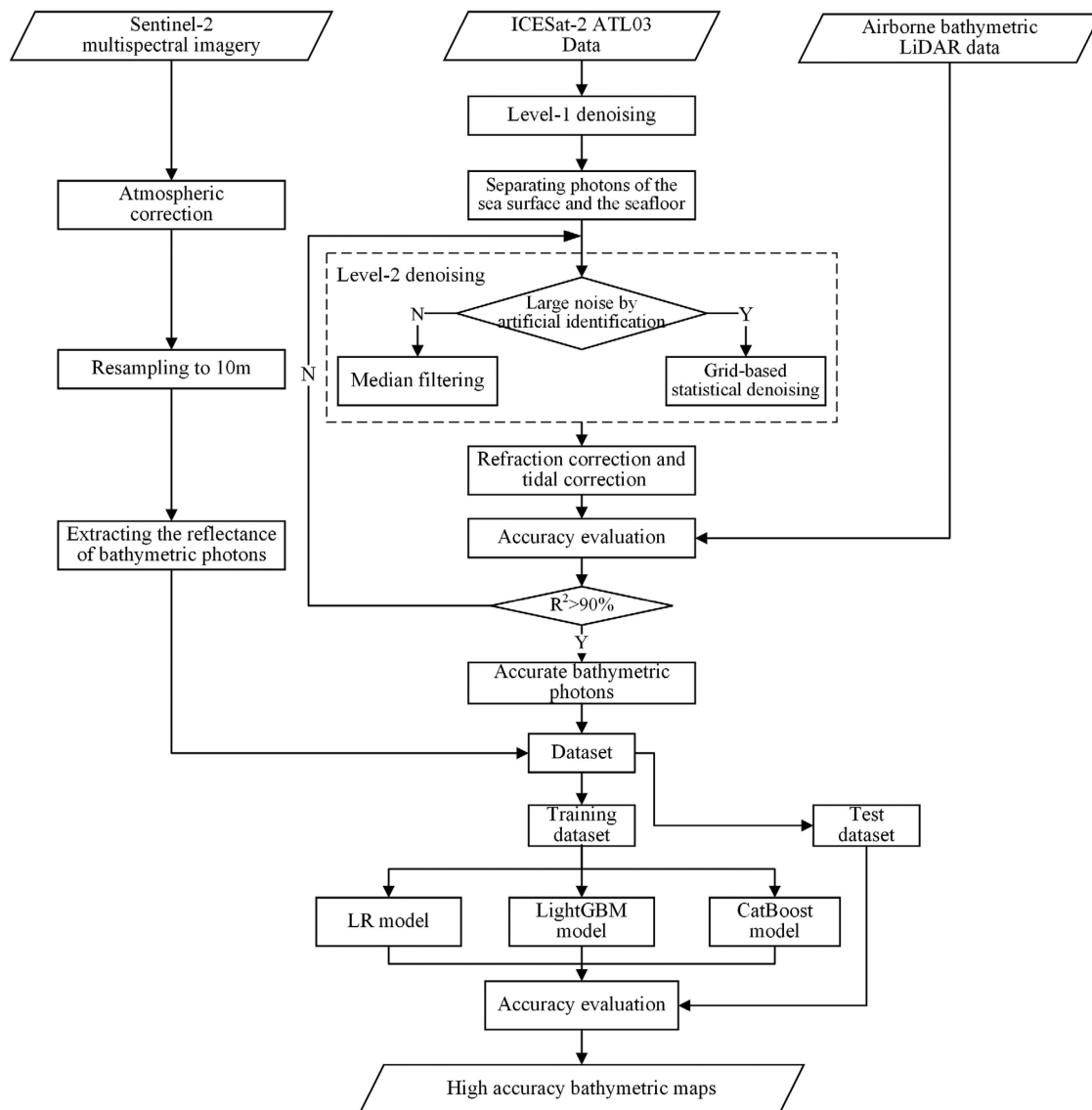


Fig. 1. Flowchart of this study.

A modified density-based spatial clustering of applications with noise method was used to extract signal photons and noise photons from ATL03 data and apply them to SDB [25]. The multiplex Gaussian fitting method was used to extract the sea surface photons from ATL03 data, and then the final photons were obtained after three median filtering iterations [26]. The water signal was extracted directly from the ATL03 data file using a semiautomatic method, which is effective in clear water and good weather conditions [22]. The adaptive variable ellipse filtering bathymetric method was proposed to accurately sort the photons at the bottom of the sea to obtain bathymetric data [27]. The difficulty of signal photon extraction is increased by the influence of environmental factors and a nonuniform noise distribution. The existing denoising algorithms still cannot eliminate manual and empirical threshold selection. Hence, a simple, feasible, and accurate process to extract accurate bathymetric photons is necessary.

In this study, we first extract accurate bathymetric photons from ICESat-2 ATL03 data, and then the following steps are completed:

- 1) removing noise photons with large errors by Level-1 denoising,
- 2) separating the sea surface and seafloor photons,
- 3) obtaining accurate bathymetric photons by Level-2 denoising.

The final photons replace in situ bathymetric data combined with Sentinel-2 imagery to build ML methods to obtain bathymetric maps of Waimanalo and the Antelope Reef. Simultaneously, the accuracy index of the methods is evaluated and analyzed. We can combine the benefits of both types of data and perform high-accuracy shallow water bathymetry on a massive scale by fusing the ICESat-2 ATL03 data with multispectral satellite imagery. The framework of this study is shown in Fig. 1.

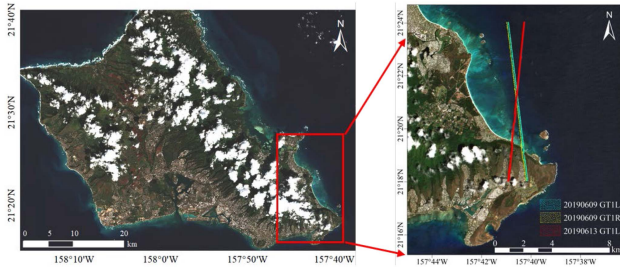


Fig. 2. Waimanalo and ICESat-2 ATL03 data shown in Sentinel-2 true-color.

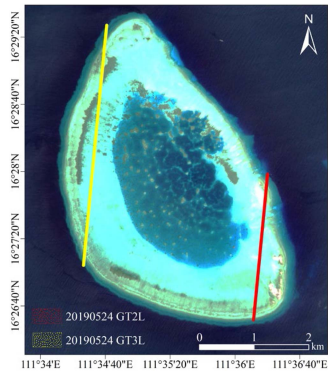


Fig. 3. Antelope Reef and ICESat-2 ATL03 data shown in Sentinel-2 true-color.

II. MATERIALS

A. Study Areas

In this article, there are two study areas. The first study area is Waimanalo (see Fig. 2), which is located to the southeast of Oahu. Oahu is the third largest Hawaiian island, covers an area of approximately 1546 km², and is the most populated island. The island has two parallel mountain ranges, the Koolau and the Waianae ranges, which are connected by a central plateau. The coast is tortuous, and there are many coral reefs along it. There is also clear water, making this area suitable for SDB.

The second study area is Antelope Reef, which is situated in the southwest corner of the Yongle Atoll in the Paracel Islands, as shown in Fig. 3. This atoll becomes a closed lagoon inside after low tide.

B. Data Collection

1) *Sentinel-2 Multispectral Imagery*: The Sentinel-2 satellites, which were launched on 23 June 2015 and 7 March 2017, respectively, are comprised of Sentinel-2A and Sentinel-2B. They were placed in the same sun-synchronous orbit with a 180° phase shift. Each satellite carries a multispectral instrument and orbits the Earth at an altitude of 786 km. The spectral range covers the visible light, near-infrared, and shortwave infrared regions, and there are three spectral bands in the red-edge range, which have high resolution and can be used to extract more water information hidden in the band. There are a total of 13 spectral bands, and they have distinct spatial resolutions at the ground that range from 10 to 60 m [28].

TABLE I
INFORMATION ON THE STUDY AREAS AND THE DATA USED IN DETAIL

Data/Study Areas	Waimanalo	Antelope Reef
Latitude	21. 31°–21. 38°N	16. 44°–16. 50°N
Longitude	157. 71°–157. 64°W	111. 57°–111. 62°E
Sentinel-2	16 April 2019	1 March 2019
Multispectral Imagery	11 May 2019	18 July 2020
	10 June 2019	25 November 2020
ICESat-2	2019/06/09 GT1L	2019/05/24 GT2L
ATL03 Data	2019/06/09 GT1R	2019/05/24 GT3L
	2019/06/13 GT1L	/
In situ Measurements	Airborne Bathymetric LiDAR Data (SHOALS)	/

The Sentinel-2 Level-1C product used in this study is free and available from the European Space Agency (<https://scihub.copernicus.eu/dhus/#/home>). The Level-1C product is transformed into the Level-2A product using the Sen2Cor toolbox, followed by resampling to 10 m in the free Sentinel Application Platform.

To avoid errors caused by environmental conditions such as cloud shadow, sun glint, and waves when imaging, three Sentinel-2 Level-1C products are used in each study area. Detailed information on the two study areas and the data used are shown in Table I.

2) *ICESat-2 ATL03 Data*: In this study, the ICESat-2 Level-2 ATL03 data, Version 5 (<https://search.earthdata.nasa.gov/search>), which include the latitude, longitude, height, and along-track distance (along with the associated uncertainties) of each telemetered photon event (lat_ph, lon_ph, h_ph, dist_ph_along, and ref_elev), are used [29].

3) *Airborne Bathymetric LiDAR Data*: The bathymetric data for Waimanalo were collected from the Scanning Hydrographic Operational Airborne LiDAR Survey (SHOALS) system and are available at <http://www.soest.hawaii.edu/coasts/data/oahu/shoals.html>. The SHOALS system consists of near-infrared and green lasers, a laser oscillator, an optical receiver, and a data processing unit and is capable of detecting depths up to 40 m. The error in the vertical direction is less than 0.15 m, and the point spacing between the sounding data is between 3 and 15 m [30]. It is a relatively mature airborne laser bathymetry system. These data are used to validate the accuracy of denoising and SDB on Waimanalo.

C. ATL03 Data Denoising and Accurate Bathymetric Photon Extraction

1) *Level-1 Denoising*: The purpose of Level-1 denoising is the use of a frequency histogram to remove obvious noise photons from the raw data.

Based on the three laser beams focused on Waimanalo [see Figs. 5(a) and 7(a)], there are very large elevation distribution differences and obvious noise photons. Moreover, the data require Level-1 denoising to control the elevation range of the

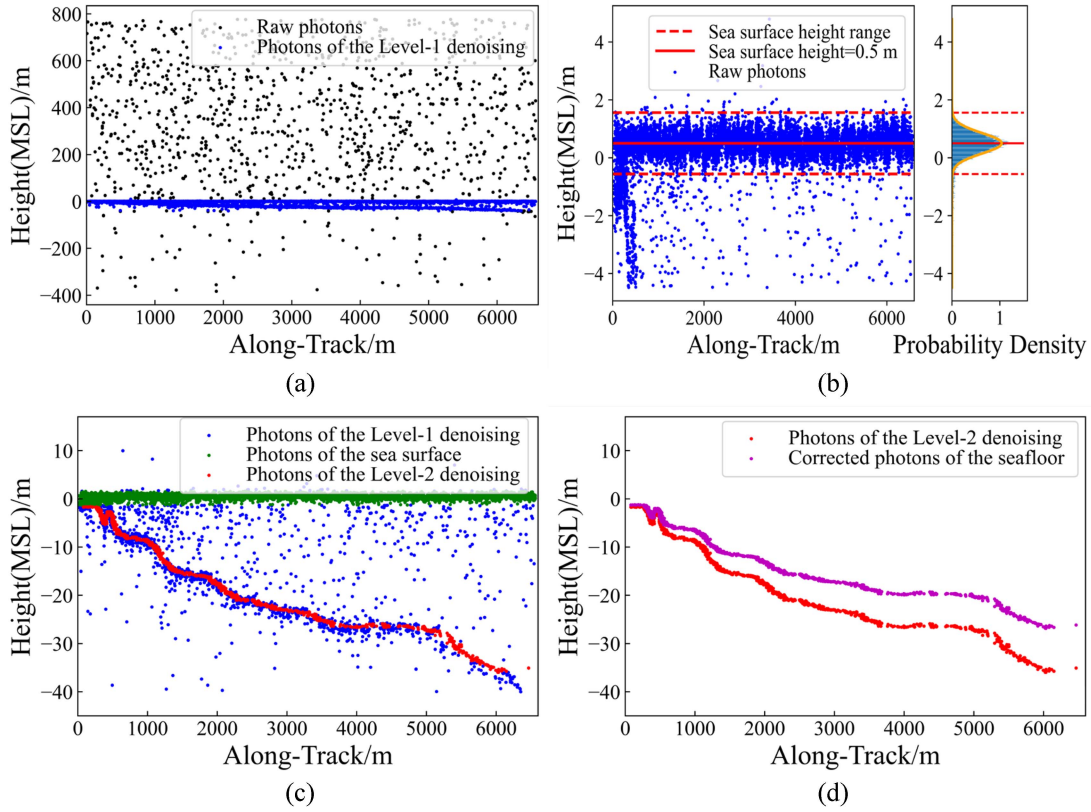


Fig. 5. Flow of 20190609 GTIL ICESat-2 ATL03 data processing based on median filtering.

$$\alpha = \sin^{-1} \left(\frac{B \sin \phi}{C} \right). \quad (6)$$

Finally

$$\Delta Y = C \cos \beta \quad (7)$$

$$\Delta Z = C \sin \beta. \quad (8)$$

Tidal correction is performed by subtracting the tidal data at the time of photon acquisition from the bathymetric value of the bathymetric photon and adding the tidal data at the time of Sentinel-2 imaging.

III. METHODS

A. ML Models

The models used in this article are as follows, and all are performed on Python 3.9.7.

1) *Linear Regression Model*: Linear regression (LR) is a model for prediction that is based on the linear combination of independent variables. In SDB, there are multiple independent variables, so the LR method is used. This method is easy to understand and apply, and the weights can visually express the importance of each band

$$y = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n + b \quad (9)$$

where y is the dependent variable, x_1, x_2, \dots, x_n are the independent variables (model inputs), i.e., the radiometric brightness

value, $\theta_1, \theta_2, \dots, \theta_n$ are the regression coefficients of the independent variables x_i and b is the intercept.

2) *Light Gradient Boosting Machine Model*: Light gradient boosting machine (LightGBM) is a model based on gradient boosting decision tree (GBDT) ([33]).

It uses a histogram-based partitioning algorithm instead of the traditional presorted traversal algorithm, which has faster parallel training efficiency, lower memory consumption, and higher accuracy. In addition, it has the advantages of supporting various distributions, handling massive volumes of data more adaptively, and effectively preventing overfitting. LightGBM not only increases the accuracy of the prediction but also greatly accelerates the prediction speed and reduces memory utilization [34].

3) *Categorical Boosting Model*: Yandex proposed categorical boosting (CatBoost), an innovative gradient boosting technology [35], [36]. This method has undergone significant parallelization enhancements, allowing the layout to be finished more quickly and more easily on internet networks [34].

CatBoost is an improved algorithm in the framework of the GBDT algorithm, which has the advantage of overcoming the gradient bias and effectively solving the problem of prediction bias, improving the accuracy of the algorithm, enhancing the generalization ability, and preventing the occurrence of overfitting.

In this study, the reflectance data from all bands of the multispectral images are used as input for each model, whereas

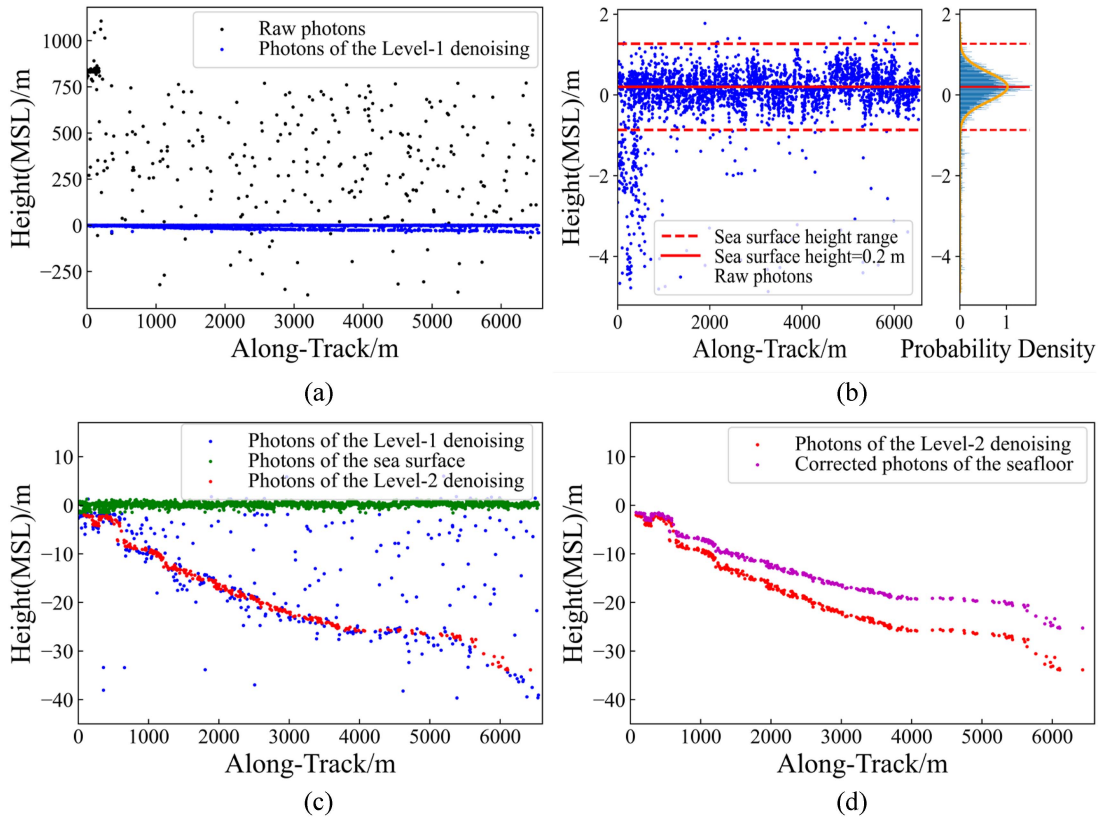


Fig. 6. Flow of 20190609 GT1R ICESat-2 ATL03 data processing based on median filtering.

the output of the model is the extracted bathymetric data from accurate bathymetric photons; 80% of the dataset is used as the training set. The accuracy of the SDB is verified using airborne bathymetric LiDAR data for the Waimanalo area. For Antelope Reef data, where no in situ bathymetry data are available, the remaining 20% of the ICESat-2 ATL03 data are utilized for verification.

B. Accuracy Evaluation Methods

The denoising process is evaluated using R^2 (coefficient of determination), root mean square error (RMSE), mean absolute error (MAE), and mean relative error (MRE). The RMSE, MAE, and MRE are used to evaluate the models

$$R^2 = 1 - \frac{\sum_{i=1}^n (h_i - \hat{h}_i)^2}{\sum_{i=1}^n (h_i - \frac{1}{n} \sum_{i=1}^n h_i)^2} \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (h_i - \hat{h}_i)^2} \quad (11)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |h_i - \hat{h}_i| \quad (12)$$

$$\text{MRE} = \frac{1}{n} \sum_{i=1}^n \left[\frac{|\hat{h}_i - h_i|}{h_i} \right] \quad (13)$$

where i is the i th bathymetric photons, n is the number of bathymetric photons used to validate the models, h_i is the bathymetric value of validation photons, and \hat{h}_i is the estimated depth.

IV. RESULTS

A. Accurately Obtaining Bathymetric Photons From ICESat-2 ATL03 Data

The algorithm described in this article is used to accurately extract bathymetric photons from data in two regions, Waimanalo and Antelope Reef, and the trajectories of the study are shown in Figs. 2 and 3. For the three laser beams (20190609 GT1L, 20190609 GT1R, and 20190613 GT1L) focused on Waimanalo, the process proposed in this article is used to filter the ATL03 data, and the feasibility of the algorithm is evaluated using the airborne bathymetric LiDAR data. Due to the lack of in situ bathymetry data on the Antelope Reef (20190524 GT2L and 20190524 GT3L), this study only analyses the denoising effect of two algorithms from a visual perspective.

Because some ATL03 data have a wide range of noise points, obvious noise points can be removed by Level-1 denoising. Figs. 5–9 show the results after Level-1 denoising, and the denoising effect is very clear. Subsequently, the elevation histogram was utilized to determine the height of the sea surface for each dataset, which is represented by the solid red line in Figs. 5(b) and 9(b). To distinguish between the photons at the sea surface and those at the seafloor, a Gaussian function was used, with photons falling within the solid red line being considered

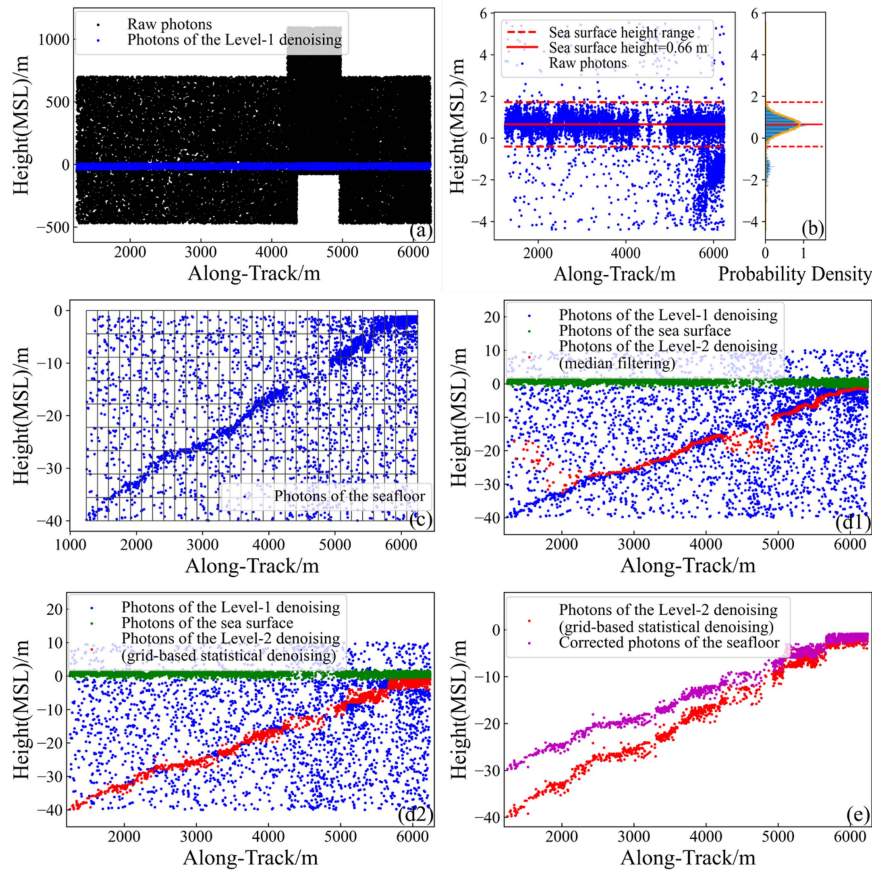


Fig. 7. Flow of 20190613 GT1L ICESat-2 ATL03 data processing based on grid-based statistical denoising.

TABLE II
ACCURACY OF ICESAT-2 ATL03 DATA DENOISING

Data	Number	R^2 (%)	RMSE (m)	MAE (m)	MRE (%)
20190609 GT1L	2506	99.85	0.53	0.36	13.49
20190609 GT1R	644	99.74	0.53	0.38	11.84
20190613 GT1L median filtering	4467	93.38	2.96	1.44	19.60
20190613 GT1L grid-based statistical denoising	2271	99.82	0.75	0.50	16.85

as surface photons (green points), whereas those falling below the second red dashed line were identified as seafloor photons.

For data from the 20190609 GT1L and 20190609 GT1R laser beams, median filtering is used for Level-2 denoising. In Figs. 5(d) and 6(d), the denoising results are the red points, and the refraction corrected photons are the magenta points.

In Figs. 8 and 9, d1 is the processing result based on median filtering, and d2 is the processing result based on grid-based statistical denoising. For the 20190613 GT1L data, the median filtering algorithm was not effective in accurately extracting bathymetric photons due to significant noise photons present in Fig. 7(d1), which could not be removed. Therefore, the grid-based statistical denoising algorithm proposed in this study was employed for filtering, and the results are presented in Fig. 7(d2). The grid statistics method, which involves dividing the grid of 17×10 (empirically specified), can retain signal photons and

eliminate noisy photons to a significant extent, as compared to the median filtering algorithm. As shown in Table II, R^2 , RMSE, MAE, and MRE improved by 6.44%, 2.21 m, 0.94 m, and 2.75%, respectively.

The results from two laser beams (20190524 GT2L and 20190524 GT3L) in the Antelope Reef area are shown in Figs. 8 and 9. The two data were divided into 100×10 and 150×10 grids [see Figs. 8(c) and 9(c)], respectively, but some noise cannot be effectively removed [see Figs. 8(d2) and 9(d2)], and signal photons are missed; the denoising effect is not as good as the result of median filtering [Figs. 8(d1) and 9(d1)]. Therefore, we still use the accurate bathymetric photon data obtained by median filtering when performing the SDB of Antelope Reef.

The extracted seafloor photons based on the suitable algorithm are significantly correlated with airborne bathymetric LiDAR data. Compared with in situ bathymetric data, as shown in

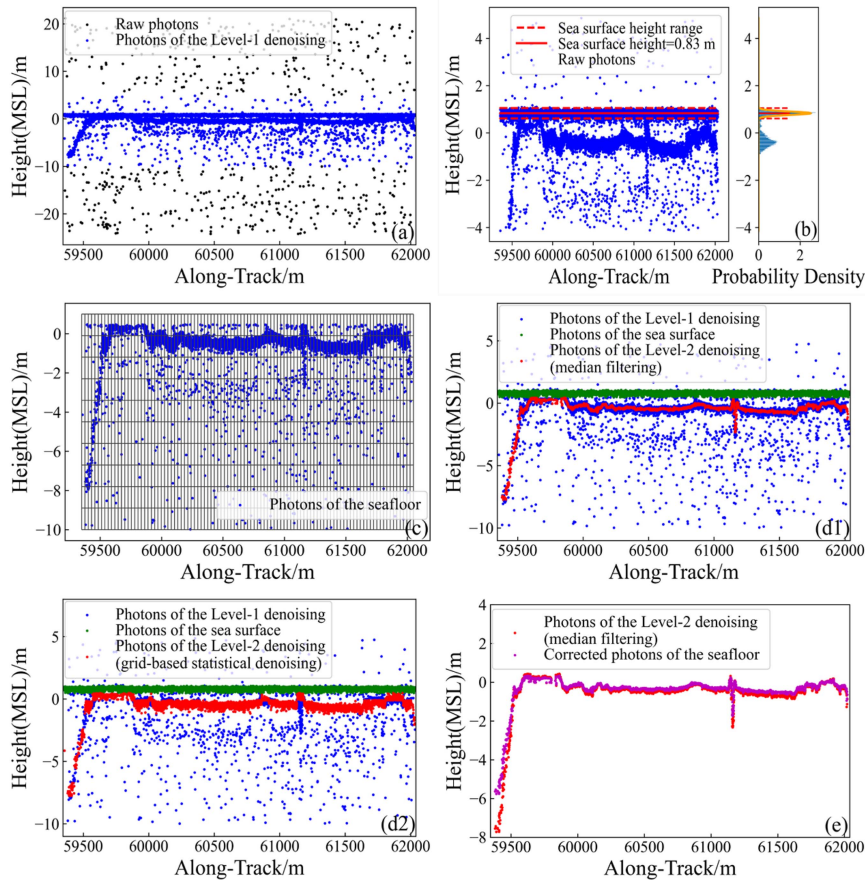


Fig. 8. Flow of 20190524 GT2L ICESat-2 ATL03 data processing based on median filtering.

Table II, the R^2 is greater than 99.74%, the RMSE is less than 0.75 m, the MAE is less than 0.50 m, and the MRE is less than 16.85%.

B. Accuracy of Different ML Models

Using accurate bathymetry photons from ICESat-2 ATL03 data and Sentinel-2 multispectral imagery, three ML methods (i.e., LR model, LightGBM model, and CatBoost model) are constructed for three dates in two study areas, Waimanalo and Antelope Reef. To assess the inversion effect, three quality indicators are used: RMSE, MAE, and MRE.

Figs. 10–14 show the error scatter diagrams of retrieved water depth and ICESat-2 ATL03 data at two areas on three dates. The black dashed line is the 1:1 line, whereas the red line corresponds to the regression line. N is the number of data used to validate the models. The higher the inversion effect is, the more scatters converge to the 1:1 line and the better the model fitting. The worse the inversion effect is, the more divergent the scatters are from the trend line. As seen from the figures, it can be seen that the degree of regression is lower using the LR model. The error scatter diagrams also show that the LR model has a poorer effect in deep water. Concentrated scatter distributions are observed when using the LightGBM and CatBoost models, and the trend lines are approximate to be 1:1 line, thus an ideal inversion result is obtained when using these models.

Figs. 15–19 show the shallow water bathymetric maps of Waimanalo and Antelope Reef using multivariate satellite images and ICESat-2 ATL03 data obtained by inversion using the LR, LightGBM, and CatBoost ML methods. In Figs. 15–19, panels (a), (e), and (i) show true color images of different dates in each study area. In the Waimanalo study area, it can be seen from the Sentinel-2 true color images that the water depth near the shore on the left side is shallow, the bottom can be seen, and the water depth on the right side is deeper. For the Sentinel-2 images of the Antelope Reef study area, the water depth in the surrounding area is shallow, and the water depth in the middle is deeper. Figs. 18 and 19 show the bathymetric maps of Antelope Reef obtained after inversion, and the water depth distribution is approximately 6 m. Due to the low accuracy of the LR model, the bathymetric map obtained based on the LR model has a large deviation from those obtained from the LightGBM and CatBoost models.

Tables III and IV show the accuracy index results of bathymetric inversion for the Waimanalo and Antelope Reef data. From an overall perspective, significantly improved accuracy is obtained using the LightGBM and CatBoost models compared with using the LR ML model. For the Waimanalo data, the mean RMSE, MAE, and MRE values of the inversion results using the LightGBM and CatBoost models are 0.81 m, 0.605 m, and 18.095%, respectively. For the Antelope Reef data, the average RMSE, MAE, and MRE values of the inversion results using

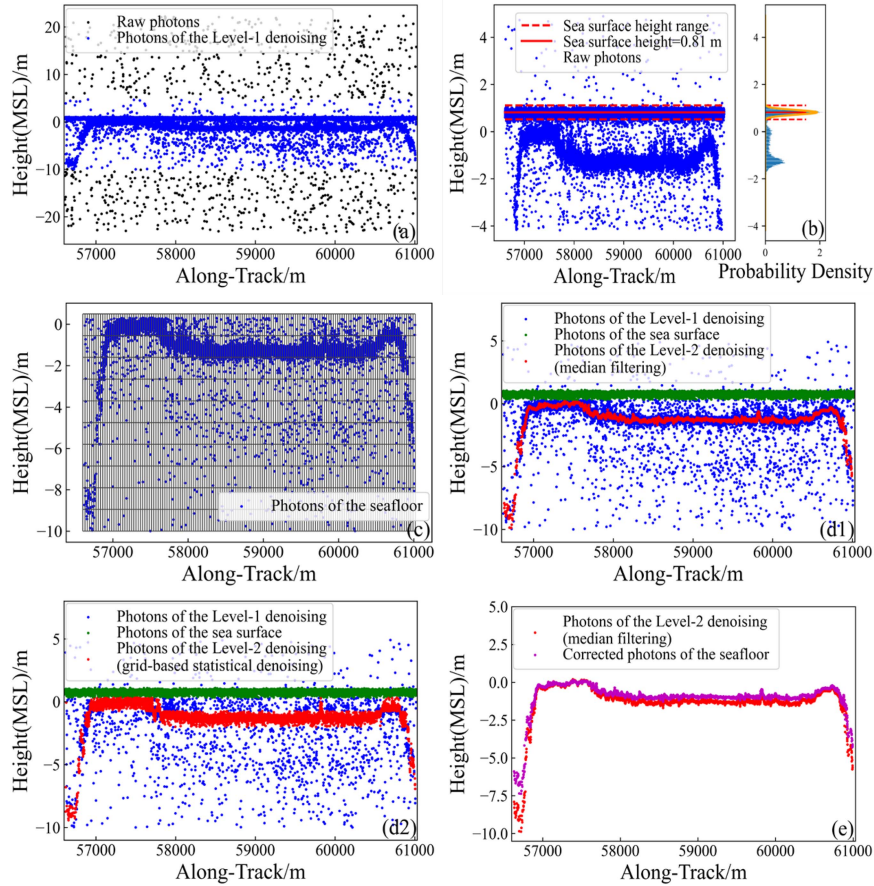


Fig. 9. Flow of 20190524 GT3L ICESat-2 ATL03 data processing based on median filtering.

the LightGBM and CatBoost models are 0.05 m, 0.02 m, and 0.945%, respectively. A positive inversion effect is observed when using the LightGBM and CatBoost models. The worst overall inversion results, with an RMSE of 2.72 m for the Waimanalo data and 0.40 m for the Antelope Reef data, are obtained using the LR model.

C. Comparison of Strong and Weak Laser Beams

The 20190609GT1L and 20190619GT1R data in the Waimanalo study area, which correspond to the pair's strong laser beam and weak laser beam, respectively, have an energy ratio of approximately 4:1. The number of photons also differs greatly. The number of strong laser beam photons is larger than the number of weak laser beam photons after denoising. Comparing the inversion results of the strong and weak beams with respect to the three indicators (see Figs. 10 and 11), better performance is obtained using the strong beam than the weak beam.

V. DISCUSSION

A. ATL03 Data Processing Procedure

As seen from Figs. 5–7, when processing ICESat-2 ATL03 data, the elevation distribution histogram is first used to remove a large number of outliers that obviously do not conform to

the elevation of the study area. For the extraction of sea surface photons, the Gaussian function adopted in this article was used to achieve the effect of sea surface extraction simply and accurately. The method of manual recognition is used to determine whether Level-2 denoising is performed by using a median filter or grid-based statistical denoising. For the laser beam in Figs. 5 and 6, a remarkable filtering effect can be obtained by using the median filter algorithm. However, for the photons shown in Fig. 7, since the data collected during the day (16:32:47 local time) are greatly affected by the solar background, it is difficult to distinguish signal photons from noise photons. Thus, the grid-based statistical denoising proposed was adopted for denoising, and the accuracy is 99%. Although not all noises were removed, a relatively obvious denoising effect was observed and the accuracy has been greatly improved. For data whose noise photons and signal photons are uniformly distributed and cannot be accurately denoised using median filtering, the grid-based statistical method can effectively achieve the denoising effect.

Based on the above-mentioned results, we can prove the effectiveness and feasibility of the proposed method. Therefore, the data collected from two laser beams (20190524 GT2L and 20190524 GT3L) on Antelope Reef are denoised using two methods. By trying different grid divisions, the algorithm can achieve a good denoising effect under a suitable grid. At present, the algorithm has some limitations. The divided grid size is consistent, and it is not possible to adjust the grid size

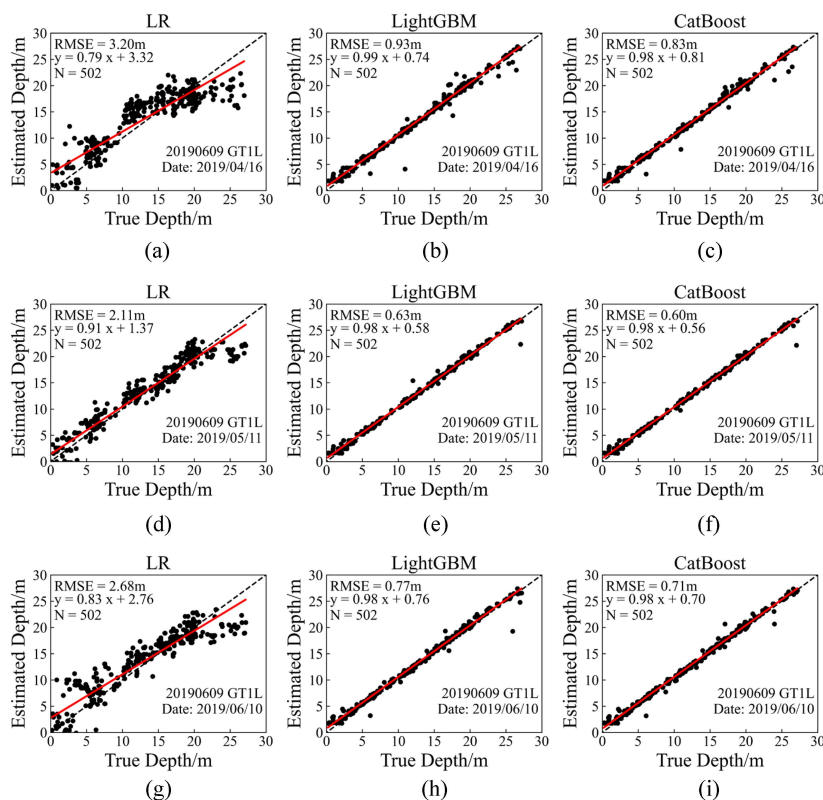


Fig. 10. Error scatter diagrams of the retrieved water depth and ICESat-2 20190609 GT1L data at Waimanalo on three dates. (a)–(c) 16 April 2019. (d)–(f) 11 May 2019. (g)–(i) 10 June 2019.

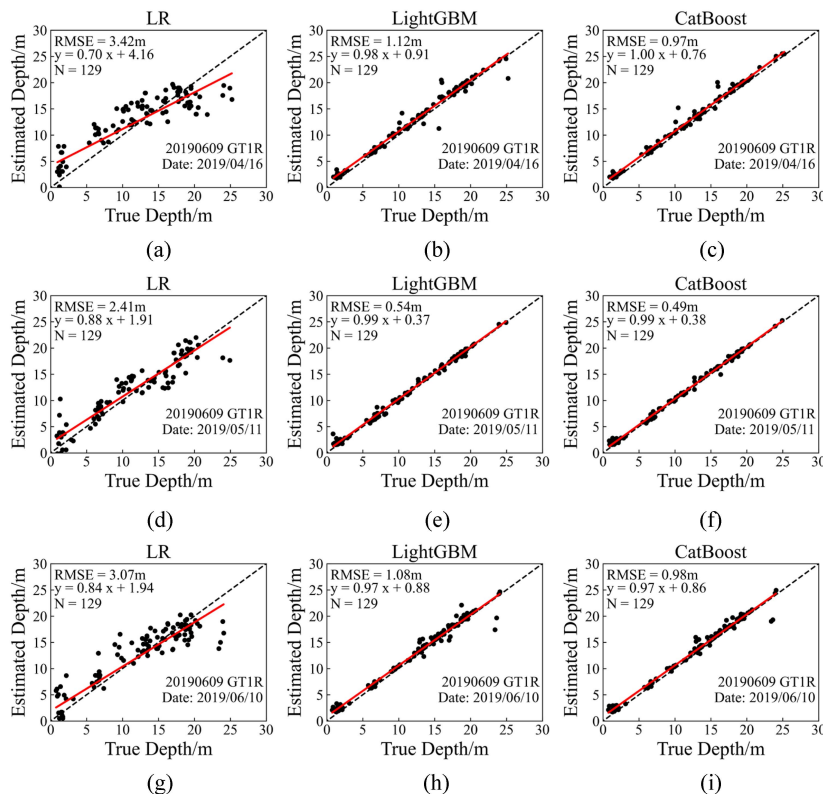


Fig. 11. Error scatter diagrams of the retrieved water depth and ICESat-2 20190609 GT1R data at Waimanalo on three dates. (a)–(c) 16 April 2019. (d)–(f) 11 May 2019. (g)–(i) 10 June 2019.

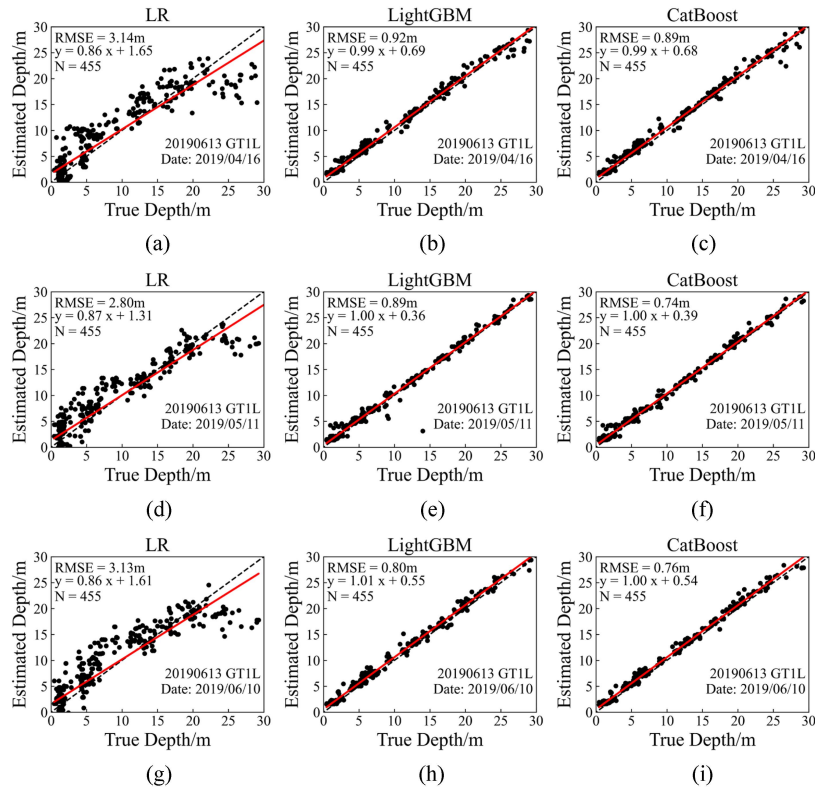


Fig. 12. Error scatter diagrams of the retrieved water depth and ICESat-2 20190613 GT1L data at Waimanalo on three dates. (a)–(c) 16 April 2019. (d)–(f) 11 May 2019. (g)–(i) On 10 June 2019.

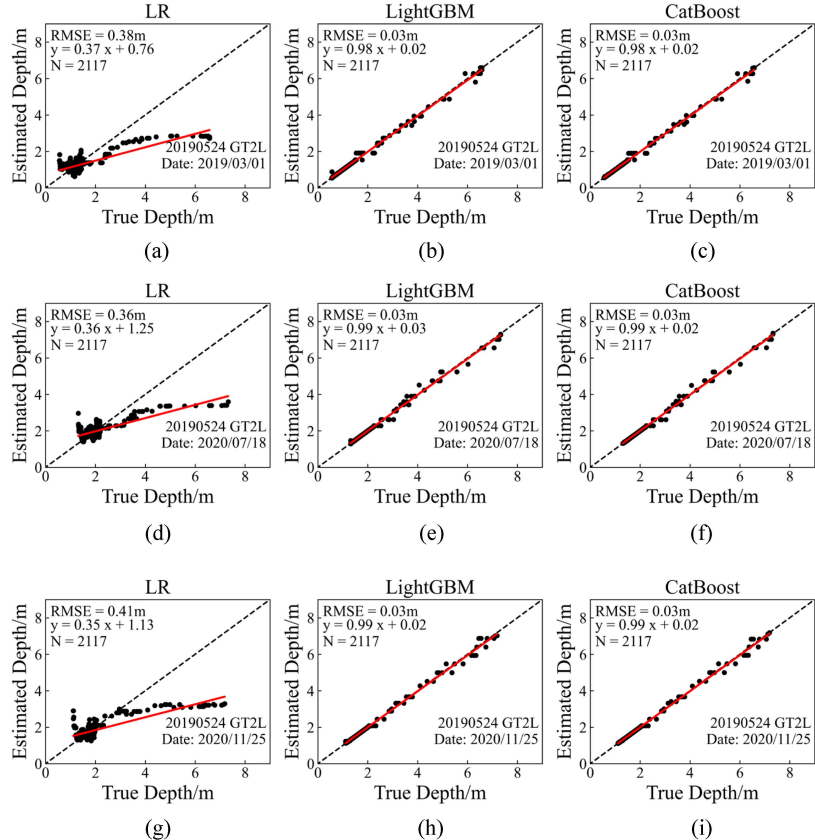


Fig. 13. Error scatter diagrams of the retrieved water depth and ICESat-2 20190524 GT2L data at Antelope Reef on three dates. (a)–(c) 1 March 2019. (d)–(f) 18 July 2020. (g)–(i) 25 November 2020.

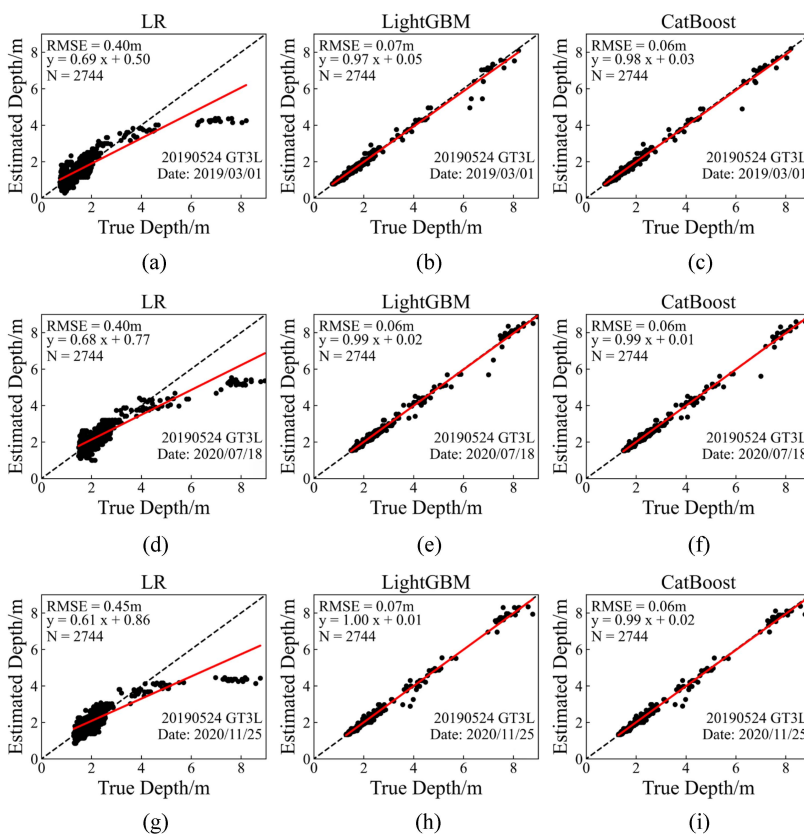


Fig. 14. Error scatter diagrams of the retrieved water depth and ICESat-2 20190524 GT3L data at Antelope Reef on three dates. (a)–(c) 1 March 2019. (d)–(f) 18 July 2020. (g)–(i) 25 November 2020.

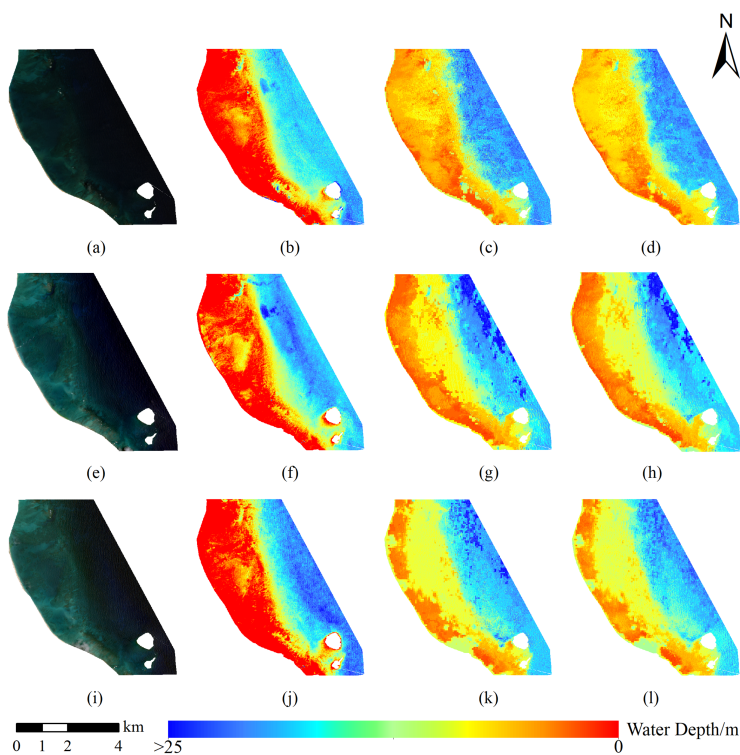


Fig. 15. Shallow water bathymetric maps derived from the three methods for Waimanalo using multirate satellite imagery and ICESat-2 20190609 GT1L data. (a) 16 April 2019. (b) LR 2019/04/16. (c) LightGBM 2019/04/16. (d) CatBoost 2019/04/16. (e) 11 May 2019. (f) LR 2019/05/11. (g) LightGBM 2019/05/11. (h) CatBoost 2019/05/11. (i) 10 June 2019. (j) LR 2019/06/10. (k) LightGBM 2019/06/10. (l) CatBoost 2019/06/10.

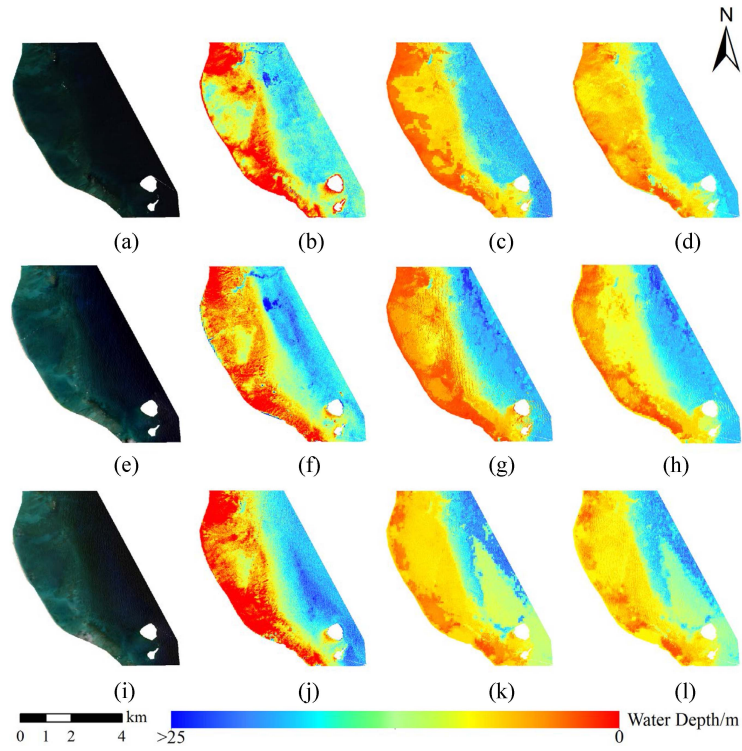


Fig. 16. Shallow water bathymetric maps derived from the three methods for Waimanalo using multidade satellite imagery and ICESat-2 20190609 GT1R data. (a) 16 April 2019. (b) LR 2019/04/16. (c) LightGBM 2019/04/16. (d) CatBoost 2019/04/16. (e) 11 May 2019. (f) LR 2019/05/11. (g) LightGBM 2019/05/11. (h) CatBoost 2019/05/11. (i) 10 June 2019. (j) LR 2019/06/10. (k) LightGBM 2019/06/10. (l) CatBoost 2019/06/10.

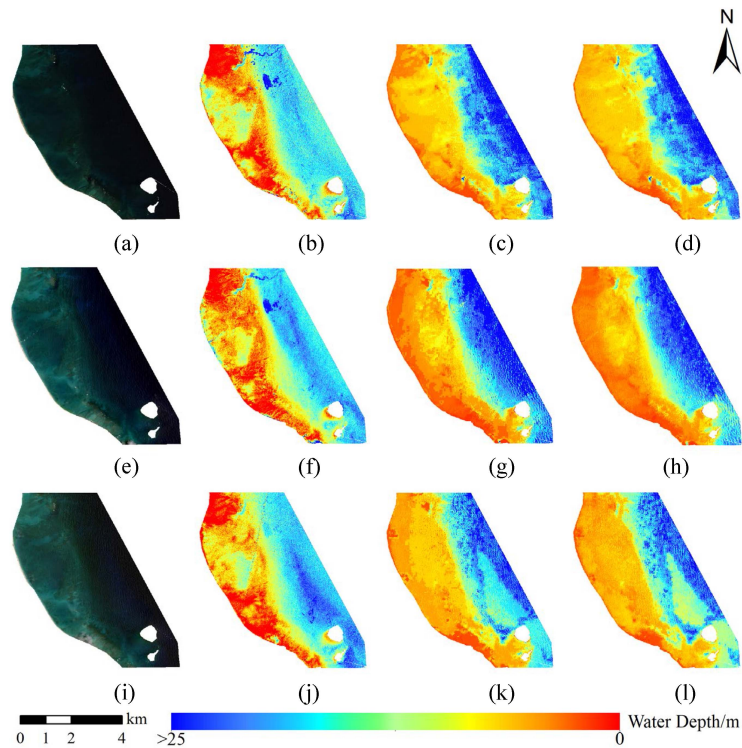


Fig. 17. Shallow water bathymetric maps derived from the three methods for Waimanalo using multidade satellite imagery and ICESat-2 20190613 GT1L data. (a) 16 April 2019. (b) LR 2019/04/16. (c) LightGBM 2019/04/16. (d) CatBoost 2019/04/16. (e) 11 May 2019. (f) LR 2019/05/11. (g) LightGBM 2019/05/11. (h) CatBoost 2019/05/11. (i) 10 June 2019. (j) LR 2019/06/10. (k) LightGBM 2019/06/10. (l) CatBoost 2019/06/10.

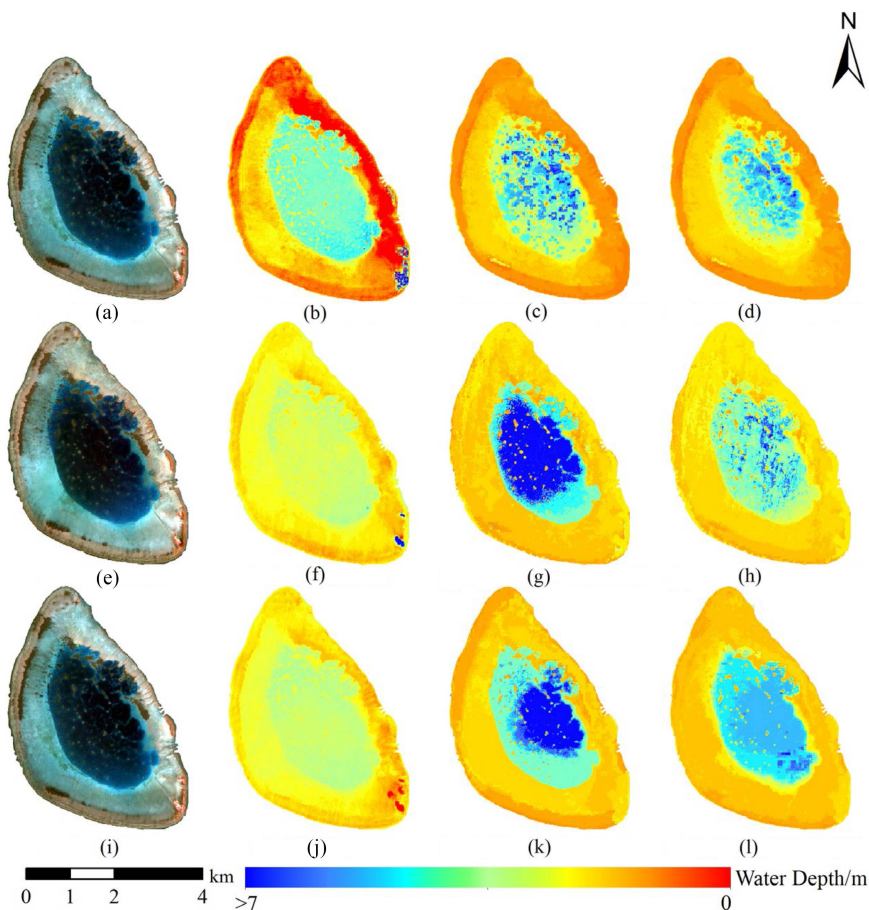


Fig. 18. Shallow water bathymetric maps derived from the three methods for Antelope Reef using multivariate satellite imagery and ICESat-2 20190524 GT2L data. (a) 1 March 2019. (b) LR 2019/03/01. (c) LightGBM 2019/03/01. (d) CatBoost 2019/03/01. (e) 18 July 2020. (f) LR 2020/07/18. (g) LightGBM 2020/07/18. (h) CatBoost 2020/07/18. (i) 25 November 2020. (j) LR 2020/11/25. (k) LightGBM 2020/11/25. (l) CatBoost 2020/11/25.

for different data. For example, obvious signal photons were missing in Fig. 8(d2), which is the area where the algorithm needs to be improved in the future. However, on the whole, the method based on grid-based statistical denoising is comparable to the median filtering, which is a wide range of algorithms, and its applicable data are more extensive, with strong universality.

Consequently, in both study areas, the denoising process proposed has excellent performance in terms of accurately obtaining bathymetric photons from daytime and nighttime ATL03 data, which is crucial for using ICESat-2 ATL03 data for SDB.

B. SDB With ICESat-2 Photons

To compare and analyze the accuracy of the three models, scatter plots of the errors of the three models for the two study areas and three dates are created, and the RMSE, MAE, and MRE values were calculated. For the three multispectral images used in the same study area, the depth inversion results were affected and biased to some extent due to the different imaging moments of Sentinel-2, which are partially affected by clouds and shadows. In analyzing the inversion results for the three dates, the errors on 16 April 2019 are relatively large at Waimanalo, which may be caused by several obvious clouds

in the image and errors in tidal correction [25]. However, the results are generally consistent, mainly because the study areas are more remote and less affected by human activities, and the selected images are almost cloud-free. Therefore, there is no effect on the inversion method, and it is more reliable to take the average value as the accuracy analysis of the bathymetric inversion methods. Overall, the RMSEs of both the LightGBM and CatBoost models are less than 1.12 m, the MAEs are less than 0.82 m, and the MREs are less than 27.01%. The RMSE, MAE, and MRE values of the LR model are less than 3.42 m, 2.90 m, and 93.10%, respectively. Very significant inversion effects are observed using the LightGBM and CatBoost models compared with those observed using the LR model. It can be concluded that the nonlinear ML method can better handle SDB in complex and diverse environments compared to the linear ML method. Fig. 10–14 are used to more directly determine the inversion effects of these three methods, and the error scatter plots also show that the LR model has difficulty accurately describing the water depth information above 20 m in the Waimanalo area. The scatter distributions of the LightGBM and CatBoost models are around the reference line, and the inversion results are more satisfactory. In addition, these models have a greater advantage in solving the bathymetric inversion problem because of the

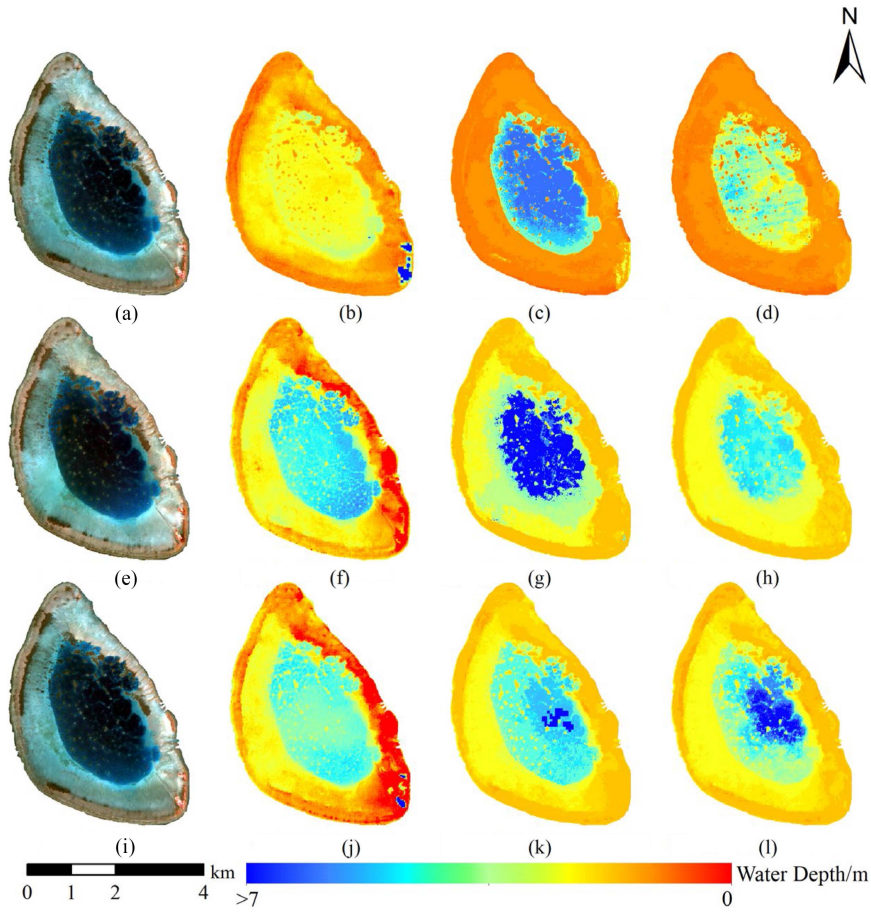


Fig. 19. Shallow water bathymetric maps derived from the three methods for Antelope Reef using multivariate satellite imagery and ICESat-2 20190524 GT3L data. (a) 1 March 2019. (b) LR 2019/03/01. (c) LightGBM 2019/03/01. (d) CatBoost 2019/03/01. (e) 18 July 2020. (f) LR 2020/07/18. (g) LightGBM 2020/07/18. (h) CatBoost 2020/07/18. (i) 25 November 2020. (j) LR 2020/11/25. (k) LightGBM 2020/11/25. (l) CatBoost 2020/11/25.

TABLE III
ACCURACY OF WATER DEPTH INVERSION AT WAIMANALO

ICESat-2 ATL03 Data	Sentinel-2 Date	LR model			LightGBM model			CatBoost model		
		RMSE (m)	MAE (m)	MRE (%)	RMSE (m)	MAE (m)	MRE (%)	RMSE (m)	MAE (m)	MRE (%)
20190609 GT1L	16 April 2019	3.20	2.61	46.98	0.93	0.70	19.53	0.83	0.67	19.56
	11 May 2019	2.11	1.61	29.97	0.63	0.41	14.07	0.60	0.39	13.90
	10 June 2019	2.68	1.97	42.61	0.77	0.60	17.41	0.71	0.57	17.18
	Mean	2.66	2.06	39.85	0.78	0.57	17.00	0.71	0.54	16.88
20190609 GT1R	16 April 2019	3.42	2.90	54.73	1.12	0.82	14.52	0.97	0.73	13.71
	11 May 2019	2.41	1.92	44.52	0.54	0.41	11.71	0.49	0.39	10.27
	10 June 2019	3.07	2.32	59.10	1.08	0.76	17.76	0.98	0.72	17.42
	Mean	2.97	2.38	52.78	0.91	0.66	14.66	0.81	0.61	13.80
20190613 GT1L	16 April 2019	3.14	2.29	76.49	0.92	0.75	27.01	0.89	0.73	26.83
	11 May 2019	2.80	2.17	93.70	0.89	0.51	18.37	0.74	0.48	18.31
	10 June 2019	3.13	2.43	87.89	0.80	0.65	24.14	0.76	0.62	23.97
	Mean	3.02	2.30	86.03	0.87	0.64	23.17	0.80	0.61	23.04
	Mean	2.88	2.25	59.55	0.85	0.62	18.28	0.77	0.59	17.91

TABLE IV
ACCURACY OF WATER DEPTH INVERSION AT ANTELOPE REEF

ICESat-2 ATL03 Data	Sentinel-2 Date	LR model			LightGBM model			CatBoost model		
		RMSE (m)	MAE (m)	MRE (%)	RMSE (m)	MAE (m)	MRE (%)	RMSE (m)	MAE (m)	MRE (%)
20190524 GT2L	1 March 2019	0.38	0.24	21.16	0.03	0.005	0.31	0.03	0.007	0.48
	18 July 2019	0.36	0.24	12.12	0.03	0.005	0.16	0.03	0.005	0.17
	25 November 2019	0.41	0.21	11.28	0.03	0.01	0.17	0.03	0.005	0.15
	Mean	0.38	0.23	14.85	0.03	0.005	0.21	0.03	0.006	0.27
20190524 GT3L	1 March 2019	0.40	0.25	16.04	0.07	0.04	2.10	0.06	0.03	2.00
	18 July 2019	0.40	0.25	10.58	0.06	0.04	1.44	0.06	0.03	1.35
	25 November 2019	0.45	0.25	11.29	0.07	0.03	1.51	0.06	0.03	1.47
	Mean	0.41	0.25	12.64	0.07	0.04	1.68	0.06	0.03	1.61
	Mean	0.40	0.24	13.75	0.05	0.02	0.95	0.05	0.02	0.94

highly complex nonlinearity between bathymetric data and the reflectivity of remote sensing images. From Figs. 15 to 19, it can be concluded that the bathymetric maps obtained based on the LR model have large deviations from those obtained using the LightGBM and CatBoost models due to the low accuracy of the LR model. The bathymetric maps based on the LightGBM and CatBoost models obtained in this study are generally consistent with the actual topography. Both the LightGBM and CatBoost models are based on the boosting algorithm, in which a weak stem of weak learners is trained with a certain combination strategy, resulting in a strong learner. LightGBM uses a leafwise algorithm with a depth limit to reduce error; thus, better accuracy can be obtained using the same number of splits. In CatBoost, the need for extensive hyperparameter optimization is reduced, which subsequently reduces the probability of overfitting. Therefore, higher inversion accuracy can be obtained by applying these two models to bathymetric inversion problems.

Based on the results of analyzing the strong and weak beams, it can be concluded that the inversion effect is better if there are more data points simulated in the same study area when the depth distribution is approximately the same. When establishing the bathymetric inversion method, more samples of a priori bathymetric data are beneficial to obtain a better training model. Thus, a strong laser beam is used for the actual bathymetric data in the Antelope Reef study area.

VI. CONCLUSION

In this study, we used ICESat-2 ATL03 data and Sentinel-2 multispectral images to build ML methods to obtain accurate bathymetric maps. This approach could be used to overcome the lack of in situ bathymetric data. Meanwhile, we developed a high-precision denoising process with a correlation of more than 99% compared to airborne bathymetric LiDAR data, as well as an accuracy improvement of 6.44%, 2.21 m, 0.94 m, and 2.75% for R^2 , RMSE, MAE, and MRE, respectively, compared to the results of median filtering. Three ML methods were constructed using the accurate bathymetric photon data obtained by this process as bathymetric data and Sentinel-2 images, and the accuracy

of SDB was verified by airborne bathymetric LiDAR data on Waimanalo. In each study area, LightGBM and CatBoost ML methods took full advantage of their ability to solve nonlinear and highly complex problems, resulting in much higher accuracy values, with mean RMSE, MAE, and MRE values of less than 0.91 m, 0.66 m, and 23.17%, respectively.

Comparing our research to the previous studies mentioned above, it can be noted that the biggest improvement is the use of ICESat-2 ATL03 data and the ability to accurately obtain bathymetric photons without many parameters according to the denoising process proposed in this article. Thus, SDB is no longer limited by a lack of in situ bathymetric data. The RMSE obtained by the ML model and verified with the measured water depth data is between 0.49 and 1.12 m. Our results show that the accurate bathymetric photons obtained by using the denoising process proposed in this article are fully capable of replacing the in situ bathymetric data for SDB, and it is very significant to select suitable, high-accuracy, and nonlinear ML methods for SDB. The insights gained from this study may be of assistance to map large areas and perform high-accuracy bathymetry in remote areas. In future research, we will continue to develop automatic extraction methods for signal photons.

REFERENCES

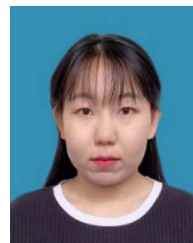
- [1] F. Eugenio, J. Marcello, and J. Martin, "High-resolution maps of bathymetry and benthic habitats in shallow-water environments using multispectral remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3539–3549, Jul. 2015.
- [2] F. C. Polcyn and D. R. Lyzenga, "Calculations of water depth from ERTS-MSS data," in *Proc. Symp. Significant Results Obtained ERTS-1, 1973*, pp. 1433–1436.
- [3] D. R. Lyzenga, "Passive remote sensing techniques for mapping water depth," *Appl. Opt.*, vol. 17, no. 3, pp. 379–383, 1978.
- [4] D. R. Lyzenga, "Shallow-water bathymetry using combined lidar and passive multispectral scanner data," *Int. J. Remote Sens.*, vol. 6, no. 1, pp. 115–125, 1985.
- [5] C. Cahalane, A. Magee, X. Monteys, G. Casal, J. Hanafin, and P. Harris, "A comparison of Landsat 8, RapidEye and Pleiades products for improving empirical predictions of satellite-derived bathymetry," *Remote Sens. Environ.*, vol. 233, Nov. 2019, Art. no. 111414.
- [6] D. R. Lyzenga, N. P. Malinas, and F. J. Tanis, "Multispectral bathymetry using a simple physically based algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2251–2259, Aug. 2006.

- [7] I. N. Figueiredo, L. Pinto, and G. Goncalves, "A modified Lyzenga's model for multispectral bathymetry using Tikhonov regularization," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 53–57, Jan. 2016.
- [8] J. M. Paredes and R. E. Spero, "Water depth mapping from passive remote sensing data under a generalized ratio assumption," *Appl. Opt.*, vol. 22, no. 8, pp. 1134–1135, 1983.
- [9] J. Wei, D. L. Civco, and W. C. Kennard, "Satellite remote bathymetry: A new mechanism for modelling," *Photogrammetric Eng. Remote Sens.*, vol. 58, no. 5, pp. 545–549, 1992.
- [10] A. H. Benny and G. J. Dawson, "Satellite imagery as an aid to bathymetric charting in the red sea," *Cartographic J.*, vol. 20, no. 1, pp. 5–16, 2013.
- [11] J. Liang, J. Zhang, Y. Ma, and C.-Y. Zhang, "Derivation of bathymetry from high-resolution optical satellite imagery and USV sounding data," *Mar. Geodesy*, vol. 40, no. 6, pp. 466–479, 2017.
- [12] J. C. Sandidge and R. J. Holyer, "Coastal bathymetry from hyperspectral observations of water radiance," *Remote Sens. Environ.*, vol. 65, pp. 341–352, 1988.
- [13] M. D. M. Manessa, M. Haidar, M. Hastuti, and D. K. Kresnawati, "Determination of the best methodology for bathymetry mapping using spot 6 imagery: A study of 12 empirical algorithms," *Int. J. Remote Sens. Earth Sci.*, vol. 14, no. 2, pp. 127–136, 2017.
- [14] J. Zhang, S. Li, and M. Wang, "Water depth inversion based on Landsat-8 data and random forest algorithm," *J. Phys., Conf. Ser.*, vol. 1437, no. 1, 2020, Art. no. 012073.
- [15] J. Zhang, Y. Zhu, X. Zhang, M. Ye, and J. Yang, "Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas," *J. Hydrol.*, vol. 561, pp. 918–929, 2018.
- [16] Z. Leng, J. Zhang, Y. Ma, and J. Zhang, "Underwater topography inversion in liaodong shoal based on GRU deep learning model," *Remote Sens.*, vol. 12, no. 24, 2020, Art. no. 4068.
- [17] Y. Zhao, L. Zhao, H. Zhang, and B. Fu, "LSTM-based remote sensing inversion of largescale sand wave topography of the Taiwan banks," *Remote Sens.*, vol. 13, no. 16, 2021, Art. no. 3313.
- [18] J. Zhu et al., "Shallow water bathymetry retrieval by optical remote sensing based on depth-invariant index and location features," *Can. J. Remote Sens.*, vol. 48, no. 4, pp. 534–550, 2022.
- [19] A. Misra, Z. Vojinovic, B. Ramakrishnan, A. Luijendijk, and R. Ranasinghe, "Shallow water bathymetry mapping using support vector machine (SVM) technique and multispectral imagery," *Int. J. Remote Sens.*, vol. 39, no. 13, pp. 4431–4450, 2018.
- [20] A. Albright and C. Glennie, "Nearshore bathymetry from fusion of Sentinel-2 and ICESat-2 observations," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 900–904, May 2021.
- [21] T. A. Neumann et al., "The ice, cloud, and land elevation satellite - 2 mission: A global geolocated photon product derived from the advanced topographic laser altimeter system," *Remote Sens. Environ.*, vol. 233, Nov. 2019, Art. no. 111325.
- [22] H. Rannald, P. S. Christiansen, P. Kliving, O. B. Andersen, and K. Nielsen, "Evaluation of a statistical approach for extracting shallow water bathymetry signals from ICESat-2 ATL03 photon data," *Remote Sens.*, vol. 13, no. 17, Sep. 2021, Art. no. 3548.
- [23] B. Cao et al., "An active-passive fusion strategy and accuracy evaluation for shallow water bathymetry based on ICESat-2 ATLAS laser point cloud and satellite remote sensing imagery," *Int. J. Remote Sens.*, vol. 42, no. 8, pp. 2783–2806, Apr. 2021.
- [24] T. Markus et al., "The ice, cloud, and land elevation satellite-2 (ICESat-2): Science requirements, concept, and implementation," *Remote Sens. Environ.*, vol. 190, pp. 260–273, 2017.
- [25] Y. Ma et al., "Satellite-derived bathymetry using the ICESat-2 lidar and sentinel-2 imagery datasets," *Remote Sens. Environ.*, vol. 250, 2020, Art. no. 112047.
- [26] H.-J. Hsu et al., "A semi-empirical scheme for bathymetric mapping in shallow water by ICESat-2 and sentinel-2: A case study in the South China Sea," *ISPRS J. Photogrammetry Remote Sens.*, vol. 178, pp. 1–19, 2021.
- [27] Y. Chen, Y. Le, D. Zhang, Y. Wang, Z. Qiu, and L. Wang, "A photon-counting LiDAR bathymetric method based on adaptive variable ellipse filtering," *Remote Sens. Environ.*, vol. 256, 2021, Art. no. 112326.
- [28] J. Liu, W. Guan, X. Wang, and J. Liu, "Optical imaging study of underwater acousto-optical fusion imaging systems," in *Proc. 13th ACM Int. Conf. Underwater Netw. Syst.*, 2018, pp. 1–5.
- [29] T. A. Neumann et al., "Cloud, and land Elevation Satellite.2 (ICESat-2) Project: Algorithm Theoretical Basis Document (ATBD) for Global Geolocated Photons (ATL03)," National Aeronautics and Space Administration, Goddard Space Flight Center, 2021.
- [30] J. L. Irish and T. E. White, "Coastal engineering applications of high-resolution lidar bathymetry," *Coastal Eng.*, vol. 35, no. 1, pp. 47–71, 1998.
- [31] B. Wang et al., "A noise removal algorithm based on adaptive elevation difference thresholding for ICESat-2 photon-counting data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 117, 2023, Art. no. 103207.
- [32] C. E. Parrish, L. A. Magruder, A. L. Neuenchwander, N. Forfinski-Sarkozi, M. Alonzo, and M. Jasinski, "Validation of ICESat-2 ATLAS bathymetry and analysis of ATLAS's bathymetric mapping performance," *Remote Sens.*, vol. 11, no. 14, 2019, Art. no. 1634.
- [33] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 3149–3157.
- [34] G. Huang et al., "Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions," *J. Hydrol.*, vol. 574, pp. 1029–1041, 2019.
- [35] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in *Proc. Neural Inf. Process. Syst.*, 2017, pp. 6639–6649.
- [36] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: Gradient boosting with categorical features support," 2018, *arXiv:1810.11363*.



Tao Xie received the M.S. degree in materials from the Guizhou University of Technology, Guiyang, China, in 2002, and the Ph.D. degree in electromagnetic fields and microwave technology from Shanghai Jiaotong University, Shanghai, in 2005.

He was a Postdoctoral Fellow in oceanography and ocean remote sensing with the Bedford Institute of Oceanography, Dartmouth, NS, Canada. He is currently a Professor with the Nanjing University of Information Science and Technology, Nanjing, China. His research interests include the modeling of electromagnetic backscattering from the ocean surface, ocean surface parameter retrieval from synthetic-aperture radar (SAR) images, oil spill detection by SAR, and Arctic sea ice remote sensing.



Ruiyao Kong received the B.S. degree in surveying and mapping engineering from the Hebei University of Technology, Tianjin, China, in 2020. She is currently working toward the M.S. degree in engineering at the Nanjing University of Information Science and Technology, Nanjing, China.

Her current research interests include machine learning and bathymetric inversion.



Abdul Nurunnabi received the B.Sc., M.Sc., and M.Phil. degrees in statistics from the University of Rajshahi, Rajshahi, Bangladesh, in 1997, 1998, and 2008, respectively, and the Ph.D. degree in spatial sciences from Curtin University, Perth, Australia, in 2015.

He is a Researcher with the University of Luxembourg, Esch-sur-Alzett, Luxembourg. His research interests include feature extraction, segmentation, classification, modeling, and estimation in image and point clouds using machine learning and deep learning.



Shuying Bai received the M.S. degree in human geography from Inner Mongolia Normal University, Hohhot, China, in 2002, and the Ph.D. degree in environmental science from the Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Beijing, China, in 2005.

She was a Postdoctoral Fellow with the Nanjing Institute of Environmental Sciences, Ministry of Environmental Protection. She is currently an Associate Professor with the Nanjing University of Information Science and Technology, Nanjing, China. Her

research interests include the application of remote sensing and GIS technology in meteorological disasters, ecological environments, carbon sources, and carbon sinks.



Xuehong Zhang received the M.S. degree in remote sensing of resources and environment from Beijing Normal University, Beijing, China, in 2005, and the Ph.D. degree in geographic information science from Nanjing University, Nanjing, China, in 2012.

He was a Postdoctoral Fellow with the Nanjing University of Information Science and Technology, Nanjing, China, where he is currently an Associate Professor. His research interests include agricultural meteorological disaster remote sensing and resource and environment remote sensing.